HW3

2021年10月20日 14:55

1.

(a)

$$\begin{cases} C = (X'X)^{-1}X' \\ X = \begin{bmatrix} 1 & .85 & 54 \\ 1 & 1.121 & 49.6 \\ \vdots & \vdots & \vdots \\ 1 & .61 & 64.5 \end{bmatrix} \end{cases} \Rightarrow C = \begin{bmatrix} -0.035 & 0.295 & \cdots & -1.020 \\ -0.119 & -0.049 & \cdots & -0.182 \\ 0.006 & -0.003 & \cdots & 0.026 \end{bmatrix}$$

$$\begin{cases} B = CY \\ Y = \begin{bmatrix} \ln(1352) \\ \ln(279) \\ \vdots \\ \ln(59) \end{bmatrix} \end{cases} \Rightarrow B = \begin{bmatrix} 7.978 \\ -1.127 \\ -0.025 \end{bmatrix}$$

Thus the estimated regression is

$$\hat{Y} = 7.978 - 1.127x_1 - 0.025x_2$$

(b)

$$\begin{cases} SS_R = \sum_{i=1}^{n} r_i^2 = r'r \\ r = Y - XB \\ r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \end{cases} \Rightarrow \begin{aligned} SS_R &= Y'Y - Y'XB \\ &= Y'Y - B'X'Y \\ &\approx 22.149 \end{aligned}$$

So the estimate of the variance of the error term is 22.149.

2.

(a) $SS = \sum_{i=1}^{n} (Y_i - \hat{\beta}x_i)^2$

$\because$ the expected response corresponding to the $x=0$ is 0

$\therefore E(e) = 0$

$$\frac{\partial SS}{\partial \beta} = -2\sum_{i=1}^{n} x_i(Y_i - \hat{\beta}x_i)$$

Setting it equal to 0, then

$$\sum_{i=1}^{n} x_iY_i = \hat{\beta}\sum_{i=1}^{n} x_i^2$$

$$\hat{\beta} = \frac{\sum x_iY_i}{\sum_i x_i^2}$$

(b) Y_i is the response corresponding to the input value x_i, then $Y_1, Y_2, \ldots, Y_n$ are independent and $Y_i \sim N(\beta x_i, \sigma^2)$

Since $\hat{\beta} = \dfrac{\sum_i x_iY_i}{\sum_i x_i^2}$

then $E(\hat{\beta}) = \dfrac{\sum_i x_i E(Y_i)}{\sum_i x_i^2} = \beta$

$$Var(\hat{\beta}) = \frac{Var(\sum_i x_iY_i)}{(\sum_i x_i^2)^2} = \frac{\sum_i x_i Var(Y_i)}{(\sum_i x_i^2)^2} = \frac{\sigma^2}{\sum_i x_i^2}$$

(c) $SS_E = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_i (Y_i - \hat{\beta}x_i)^2$

(equivalently, SS_R)

$$\begin{aligned} &= \sum_i Y_i^2 - \hat{\beta}\sum x_iY_i \\ &= \sum_i Y_i^2 - \frac{\sum x_iY_i}{\sum_i x_i^2}\sum x_iY_i \\ &= \sum_i Y_i^2 - \frac{(\sum x_iY_i)^2}{\sum_i x_i^2} \end{aligned}$$

(d) To test $H_0: \beta = \beta_0$ versus $H_a: \beta \neq \beta_0$

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_i x_i^2}$$

$$\frac{\hat{\beta} - \beta}{\sqrt{\sigma^2 / \sum x_i^2}} = \sqrt{\sum_i x_i^2}\,\frac{\hat{\beta} - \beta}{\sigma} \sim N(0,1)$$

$$\frac{\hat\beta-\beta}{\sqrt{\sigma^2/\sum_i x_i^2}}=\sqrt{\sum_i x_i^2}\,\frac{\hat\beta-\beta}{\sigma}\sim N(0,1)$$

and is independent of $\frac{SSE}{\sigma^2}\sim\chi^2_{n-1}$

Hence, it follows that $\frac{\sqrt{\sum_i x_i^2}(\hat\beta-\beta_0)/\sigma}{\sqrt{\frac{SSE}{\sigma^2(n-1)}}}=\sqrt{\frac{\sum_i x_i^2(n-1)}{SSE}}(\hat\beta-\beta)\sim t_{n-1}$

Therefore if H_0 is true (and so $\beta=\beta_0$), then

$$\sqrt{\frac{\sum_i x_i^2(n-1)}{SSE}}(\hat\beta-\beta_0)\sim t_{n-1}$$

A significant level γ test of H_0 is to

reject H_0 if $\sqrt{\frac{\sum_i x_i^2(n-1)}{SSE}}|\hat\beta-\beta_0|>t_{\gamma/2,n-1}$

accept H_0 otherwise

$$v=\sqrt{\frac{\sum x_i^2(n-1)}{SSE}}|\hat\beta-\beta_0|$$

and then rejecting H_0 if the desired significant level is at least as large as

$$p\text{-value}=P\{|T_{n-1}|>v\}$$
$$=2P\{T_{n-1}>v\}$$

3. (a) Given data: $\{(x_i,y_i)\}$

x_i = Anger Score $\in[25,81]$

$y_i=\begin{cases}0 & \text{didn't have a second heart attack}\\ 1 & \text{Had a second heart attack}\end{cases}$

Model: $\hat y_i=\frac{e^{-\beta^T x}}{1+e^{-\beta^T x}}+e_i$ = Prob (have 2nd heart attack within 5 years)

$(1-h_\beta(x_i))$ where $-\beta^T x_i=-\beta_0-\beta_1 x_i$

(b)

```
>> X=[81;76;70;69;63;60;51;45;40;36;29;25];Y=[1;1;0;1;0;1;1;0;1;0;0;1];
>> beta = glmfit(X,Y)

beta =

    0.2272
    0.0066
```

Our estimate parameters $\hat\beta\approx\begin{bmatrix}0.2272\\0.0066\end{bmatrix}$

(c)

$$\hat y_i=\frac{e^{(0.2272+0.0066\times 55)}}{1+e^{(0.2272+0.0066\times 55)}}$$
$$\approx 0.643$$

4. $$\sum_{i=1}^n\alpha_i f(x_i)\le f\left(\sum_{i=1}^n\alpha_i x_i\right)$$

f is a concave function, $0<\alpha_i<1$, $\sum_i\alpha_i=1$

$$\Downarrow$$

$$f(\alpha x+(1-\alpha)y)\ge\alpha f(x)+(1-\alpha)f(y)\quad(4.1)$$

① when $n=1$, $\sum_{i=1}^n\alpha_i f(x_i)=\alpha_1 f(x_1)=f(x_1)$

$\sum_i\alpha_i=1\Rightarrow\alpha_1=1$

$f\left(\sum_{i=1}^n\alpha_i x_i\right)=f(\alpha_1 x_1)=f(x_1)$

① when $n=1$, $\sum_{i=1}\alpha_i f(x_i) = \alpha_1 f(x_1) = f(x_1)$

$\sum_i \alpha_i = 1 \Rightarrow \alpha_1 = 1$

$f(\sum_{i=1} \alpha_i x_i) = f(\alpha_1 x_1) = f(x_1)$

then the inequality holds when $n=1$.

② when $n=2$, $\sum_i \alpha_i = \alpha_1 + \alpha_2 = 1$

$\sum_i \alpha_i f(x_i) = \alpha_1 f(x_1) + \alpha_2 f(x_2) = \alpha_1 f(x_1) + (1-\alpha_1) f(x_2)$

$f(\sum_i \alpha_i x_i) = f(\alpha_1 x_1 + \alpha_2 x_2) = f(\alpha_1 x_1 + (1-\alpha_1) x_2)$

Since inequality 4.1, we obtain the inequality holds when $n=2$.

③ when $n \geq 3$

$$\alpha_1 f(x_1) + \alpha_2 f(x_2) + \cdots + \alpha_n f(x_n)$$

$$= (\alpha_1+\alpha_2)\left(\frac{\alpha_1}{\alpha_1+\alpha_2} f(x_1) + \frac{\alpha_2}{\alpha_1+\alpha_2} f(x_2)\right) + \alpha_3 f(x_3) + \cdots + \alpha_n f(x_n)$$

$$\leq (\alpha_1+\alpha_2) f\left(\frac{\alpha_1}{\alpha_1+\alpha_2} x_1 + \frac{\alpha_2}{\alpha_1+\alpha_2} x_2\right) + \cdots + \alpha_n f(x_n)$$

$$\leq f\left((\alpha_1+\alpha_2)\left(\frac{\alpha_1}{\alpha_1+\alpha_2} x_1 + \frac{\alpha_2}{\alpha_1+\alpha_2} x_2\right) + \alpha_3 x_3 + \cdots + \alpha_n x_n\right)$$

$$= f(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n)$$

Summarizing, $\sum_{i=1}^{n} \alpha_i f(x_i) \leq f(\sum_{i=1}^{n} \alpha_i x_i)$

5.

$$\log\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \frac{1}{n}\log\left(\prod_{i=1}^{n} x_i\right) = \frac{1}{n}\log x_1 + \frac{1}{n}\log x_2 + \cdots + \frac{1}{n}\log x_n$$
$$= \frac{1}{n} f(x_1) + \frac{1}{n} f(x_2) + \cdots \frac{1}{n} f(x_n)$$

$$\log\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \log\left(\frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n\right)$$
$$= f\left(\frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n\right)$$

Since we have proved $\sum_{i=1}^{n} \alpha_i f(x_i) \leq f(\sum_{i=1}^{n} \alpha_i x_i)$, and $f(x) = \log x$ is concave on $(0, \infty)$, we can conclude that

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \leq \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

6.

1.

3x200 double

	1	2	3	4
1	48.0431	0	0	0
2	0	7.6402	0	0
3	0	0	2.0271e-15	0
4				
5				

(1)Singular value matrix

According to the SVD result of the data, the singular values matrix shows we can focus on the first two components.

2.

```
A_mean=mean(A);
A_subt=A-A_mean;
```

3x1 double

	1
1	21.7764
2	6.8337
3	2.2525e-15

(2)singular value column vector

covariance=1/(N-1)*A_subt*A_subt';

3x3 double

	1	2	3
1	1.0801	0.4631	1.0801
2	0.4631	0.4575	0.4631
3	1.0801	0.4631	1.0801

(3)covariance matrix

[Evec,Eval] = eig(covariance);
[~,idx]=sort(diag(Eval),'descend');
Eval=Eval(idx,idx);
Evec=Evec(:,idx);

3x1 double

	1
1	2.3830
2	0.2347
3	-8.3468e-17

(4)eigenvalues vector

3x3 double

	1	2	3
1	0.6694	0.2277	0.7071
2	0.3220	-0.9467	-6.5061e-16
3	0.6694	0.2277	-0.7071

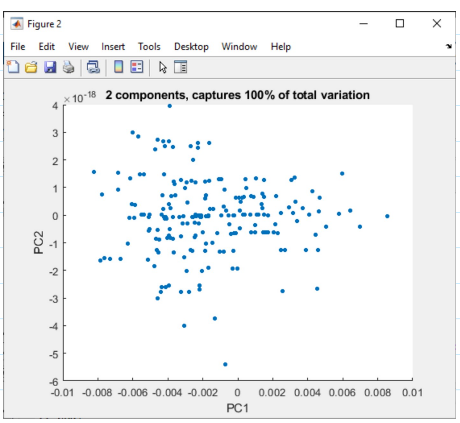
(5)the matrix whose columns are the corresponding right eigenvectors

3.

We found that after the PCA by covariance matrix, the data matrix can be approximated by two components, but by the direct SVD of de-biased and scaled data, it can be done by just one.

For the direct SVD of the de-biased, scaled data:
(plot PCA space of the first two PCs: PC1 and PC2)

```
[U,S,V]= svd(A_subt/(N-1));
sigma=diag(S);
rho = norm(sigma)^2;
C = S(1:3,1:3)*V(:,1:3)';
% first 3 coefficients for each point, same as U(:,1:3)'*A;
q = norm(sigma(1:2))^2/rho;
% part of variation captured by first 2 components
figure(2);
scatter(C(1,:),C(2,:),17,'filled')
xlabel('PC1'); ylabel('PC2');
title(sprintf('2 components, captures %.4g%% of total
variation',100*q));
```



For the PCA by covariance matrix:
(plot PCA space of the first two PCs: PC1 and PC2)

```
Eval=Eval(idx,idx);
Evec=Evec(:,idx);
Evec = Evec(:,end:-1:1);
Evec=Evec';
% generate PCA component space (PCA scores)
pc = Evec*A;
plot(pc(1,:),pc(2,:),'.')
```

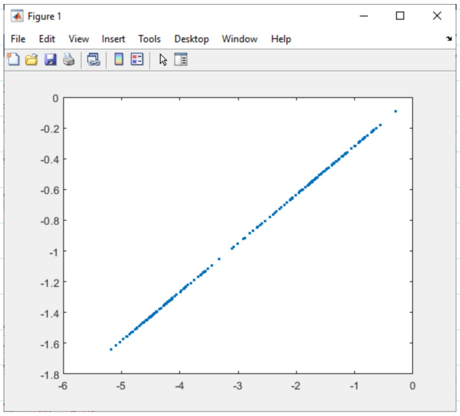


Figure 1
File Edit View Insert Tools Desktop Window Help
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
-1.8
-6
-5
-4
-3
-2
-1
0